Capturing and Treating Unobserved Heterogeneity by Response Based Segmentation in PLS Path Modeling. A Comparison of Alternative Methods by Computational Experiments

Vincenzo Esposito Vinzi, Christian M. Ringle,
Silvia Squillacciotti, Laura Trinchera

July 2007

ESSEC
CENTRE
DE RECHERCHE

Capturing and Treating Unobserved Heterogeneity by Response Based Segmentation in PLS Path Modeling. A Comparison of Alternative Methods by Computational Experiments

Vincenzo ESPOSITO VINZI
Christian M. RINGLE
Silvia SQUILLACCIOTTI
Laura TRINCHERA

Abstract:

Segmentation in PLS path modeling framework results is a critical issue in social sciences. The assumption that data is collected from a single homogeneous population is often unrealistic. Sequential clustering techniques on the manifest variables level are ineffective to account for heterogeneity in path model estimates. Three PLS path model related statistical approaches have been developed as solutions for this problem. The purpose of this paper is to present a study on sets of simulated data with different characteristics that allows a primary assessment of these methodologies.

Keywords: Partial Least Squares, Path Modeling, Unobserved Heterogeneity

Résumé :

De nos jours, les problématiques liées à la recherche d'hétérogénéité parmi les unités sont devenues critiques dans le cadre des modèles structurels PLS, notamment dans les sciences sociales. L'hypothèse de base de cette méthode, selon laquelle les données proviennent d'une population unique et homogène, s'avère souvent peu réaliste. Les techniques de classification séquentielles sur les variables manifestes sont fréquemment peu efficaces lorsque l'on veut découvrir l'hétérogénéité dans les estimations des paramètres des modèles structurels. Trois approches statistiques ont été développées comme solutions à ce problème dans le cadre des méthodes PLS. L'objectif de ce papier est de présenter une étude sur des jeux de données simulées, ayant différentes caractéristiques permettant une première évaluation des méthodes décrites. Par ces jeux de données, nous allons illustrer l'intérêt de découvrir l'hétérogénéité latente dans les applications des modèles structurels PLS, décrire les caractéristiques de chaque méthode, en comparer les points forts et les points faibles, et découvrir des aspects méthodologiques qui n'ont pas encore été traités. Ces contributions pourront aider chercheurs et praticiens à mieux comprendre les résultats parfois ambigus des modèles PLS, afin de parvenir à des conclusions analytiques plus efficaces.

Mots-clés : Partial Least Squares, Path Modeling, Unobserved Heterogeneity

JEL Classification: C39, C49

Capturing and Treating Unobserved Heterogeneity by Response Based Segmentation in PLS Path Modeling

A Comparison of Alternative Methods by Computational Experiments

Vincenzo Esposito Vinzi · **Christian M. Ringle** ·
Silvia Squillacciotti · **Laura Trinchera**

Abstract Segmentation in PLS path modeling framework results is a critical issue in social sciences. The assumption that data is collected from a single homogeneous population is often unrealistic. Sequential clustering techniques on the manifest variables level are ineffective to account for heterogeneity in path model estimates. Three PLS path model related statistical approaches have been developed as solutions for this problem. The purpose of this paper is to present a study on sets of simulated data with different characteristics that allows a primary assessment of these methodologies. We thereby illustrate the requirement to uncover unobserved heterogeneity in PLS path model applications, reveal the capabilities of each approach to deal with that issue, compare their strength and weaknesses, and provide methodological implications that have not been addressed, yet. These contributions are particularly important for researchers and practitioners to further differentiate ambiguous PLS path modeling results on the aggregate data level in order to originate effective analytical conclusions.

Keywords Partial Least Squares · Path Modeling · Segmentation · FIMIX-PLS · PLS-TPM · REBUS-PLS

Vincenzo Esposito Vinzi
ESSEC Business School,
Department of Information Systems & Decision Sciences (SID),
Avenue Bernard Hirsch, B.P. 50105,
95021 Cergy-Pontoise Cedex, France
Tel.: +33 1 34.43.36.56
Fax: +33 1 34.43.36.91
E-mail: vinzi@essec.fr

Christian M. Ringle
University of Hamburg,
Faculty of Business, Economics and Social Sciences,
Von-Melle-Park 5, 20146 Hamburg, Germany

Silvia Squillacciotti
Electricité de France, Research & Development,
1 Avenue du Général de Gaulle, 92141 Clamart, France

Laura Trinchera
University of Naples ◁ Federico II ▷,
Department of Mathematics and Statistics,
Via Cintia 27, Complesso Monte Sant'Angelo, 80126 Naples, Italy

1 Introduction

Structural Equation Modeling (SEM) and Latent Variable Path Modeling (LVP) are used to measure complex cause effect relationships (Marcoulides and Saunders, 2006; Chin, 1998). Covariance Based Structural Equation Modeling (CBSEM; Jöreskog, 1978; Rigdon, 1998) and Partial Least Squares analysis (PLS; Lohmöller, 1989; Tenenhaus et al., 2005) constitute the two corresponding statistical techniques for estimating causal models. Wold's 1982 basic PLS design or basic method of soft modeling is rather a distinctive than alternative approach compared to CBSEM (Fornell and Bookstein, 1982). Soft modeling refers to the ability of PLS to be more flexible in handling various modeling problems in situations where it is difficult or impossible to meet the hard assumptions of more traditional multivariate statistics. Within this context, "soft" is only attributed to distributional assumptions and not to the concepts, the models or the estimation techniques (Lohmöller, 1989). Applications of PLS path modeling are well established in academic literature. Recognizable examples in the management sciences discipline are Fornell et al. (1985, 1990); Gray and Meister (2004); Venkatesh and Agarwal (2006).

There are at least three critical issues that have received little or no attention in prior work. First, analyses in a PLS path modeling framework usually do not address the problem of heterogeneity. The failure to account for heterogeneity leads to ambiguous PLS path modeling results and, thus, to conclusions that are incomplete and ineffective.

Second, statistical instruments that extent and complement the PLS approach are not well developed. Especially advances towards analytical methods for clustering data have lagged behind their need in applications. A first approach is to test if data on the manifest variables level belongs to a finite number of clusters by sequential segmentation strategies such as K-means or tree clustering. These methods, however, cannot account for heterogeneity in the relationships between latent variables and are improper for forming groups of data with distinctive path model estimates (Jedidi et al., 1997). Alternatively, theoretical knowledge allows to analyze the effect of moderating factors (Henseler and Fassott, 2007). This kind of approach only accounts for heterogeneity of a limited number of relationships in the path model and is restricted to the existence of suitable moderating variables.

As a consequence of these limitations in existing methodologies, PLS path modeling requires complementary techniques for response based segmentation such as Finite Mixture Partial Least Squares (FIMIX-PLS; Hahn et al., 2002), Partial Least Squares Typological Path Modeling (PLS-TPM; Squillacciotti, 2007) and REsponse Based Units Segmetation (REBUS; Esposito Vinzi et al., 2007).

Third, knowledge is limited about the capabilities of all these methods. The systematical application of these methodologies to identify and effectively treat unobserved heterogeneity in the path model by segmentation has not been evaluated thus far. In addition, literature does not present a comparison of the methodologies and a critical analysis of their strengths and weaknesses. The failure to account for these issues entails uncertainty about the applicability of FIMIX-PLS, PLS-TPM and REBUS-PLS to analyze PLS path modeling estimates in particular research situations.

In view of these deficiencies, the primary contribution of this article is to provide an initial assessment if FIMIX-PLS, PLS-TPM and REBUS-PLS reliably identify clusters with distinctive path model relationships. We therefore simulate different kinds of data for numerical PLS path modeling examples. The simulation study allows us to evaluate the quality of results and, thus, the capabilities of each methodology under systematically changed data constellations. A comparison of the outcomes exposes strengths and weaknesses of each approach. These analyses finally substantiate the requirement and applicability of the

FIMIX-PLS, the PLS-TPM and the REBUS-PLS as analytical extensions and standard test procedures for PLS path modeling.

The findings of our study are particularly important for researchers and practitioners to confirm that PLS path modeling estimates are not affected by heterogeneity in the path model. Outcomes on the aggregate data level may otherwise be inadequate and require segmentation. FIMIX-PLS, PLS-TPM and REBUS-PLS indicate how to treat that problem by forming groups of data. Researchers and practitioners can exploit these PLS response based segmentation techniques to assure that results on the aggregate data level are not affected by unobserved heterogeneity in the path model estimates. Otherwise, FIMIX-PLS, PLS-TPM abd REBUS-PLS indicate how to treat that problem by forming groups of data. A multigroup PLS path analysis (Chin and Dibbern, 2007) on the finally established segments provides evidence if segment specific estimates are significantly different. In this case, researchers or practitioners obtain supplementary analytical results to further differentiate and render more precisely their conclusions.

The remainder of this article is organized as follows. The next section introduces FIMIX-PLS, PLS-TPM and REBUS-PLS, followed by a section about the design of this study and data simulation. Then, numerical examples allow us to demonstrate the requirement of identifying and treating heterogeneity in the path model estimates. They also provide a means of assessing and substantiating the FIMIX-PLS, the PLS-TPM and the REBUS-PLS approaches and permit an initial comparison of segmentation results. The final section concludes this article with essential implications for PLS path modeling and directions of future research.

2 Response Based Segmentation Techniques for PLS Path Modeling

2.1 Finite Mixture Partial Least Squares

2.1.1 Methodology and Algorithm

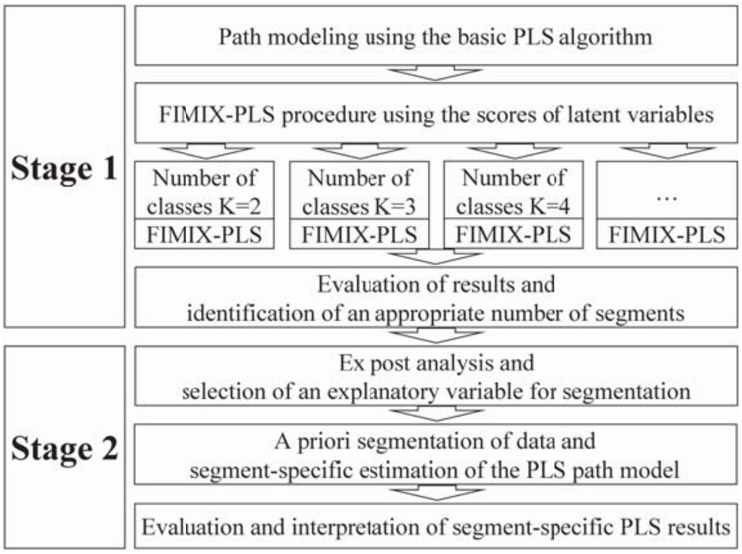
Distinctive PLS path model estimates for groups of data may be concentrated in the relationships between latent variables. FIMIX-PLS is a methodology that has been developed by Hahn et al. (2002) to capture this kind of heterogeneity for a pre-determined number of K latent classes. A comprehensive FIMIX-PLS application involves the two stages in Figure 1 (Ringle et al., 2007).

In the first stage of FIMIX-PLS, a path model is estimated by using the PLS algorithm for LVP and (empirical) data for manifest variables in the outer measurement models. The resulting scores for latent variables in the inner path model are then employed to run the FIMIX-PLS algorithm and is a key issue regarding the methodological advantages and disadvantages. On the one hand, the use of latent variable scores allows generally apply FIMIX-PLS in PLS path models (Ringle et al., 2007) regardless whether measurement models of latent variables are operationalized as formative (Diamantopoulos and Winklhofer, 2001) or reflective (Gerbing and Anderson, 1988). On the other hand, the methodology does not directly include detection of heterogeneity in the outer PLS path models.

Equation 1 expresses a modified presentation of the relationships (Table 1 provides a description of all of the symbols used in the equations presented in this paper):

$$B\eta_i + \Gamma\xi_i = \zeta_i \tag{1}$$

Fig. 1 Analytical Steps of FIMIX-PLS



Segment-specific heterogeneity of path models is concentrated in the estimated relationships between latent variables. FIMIX-PLS captures this heterogeneity and calculates the probability of each observation so that it fits into each of the pre-determined K numbers of classes. The segment-specific distributional function is defined as follows, assuming that η_i is distributed as a finite mixture of conditional multivariate normal densities $f_{i|k}(\cdot)$:

$$\eta_i \sim \sum_{k=1}^{K} \rho_k f_{i|k}(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k) \tag{2}$$

Substituting $f_{i|k}(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k)$ results in the following equation:

$$\eta_i \sim \sum_{k=1}^{K} \rho_k \left[\frac{|B_k|}{\sqrt[M]{2\pi}\sqrt{|\Psi_k|}} e^{-\frac{1}{2}(B_k\eta_i+\Gamma_k\xi_i)'\Psi_k^{-1}(B_k\eta_i+\Gamma_k\xi_i))} \right] \tag{3}$$

Equation 4 represents an EM-formulation of the log-likelihood (lnL) as the objective function for maximization:

$$lnL = \sum_i \sum_k z_{ik} ln(f(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k)) + \sum_i \sum_k z_{ik} ln(\rho_k) \tag{4}$$

An EM-formulation of the FIMIX-PLS algorithm (Figure 2) is used for statistical computations to maximize the likelihood and to ensure convergence in this model. The expectation of Equation 4 is calculated in the E-step, where z_{ik} is 1 if subject i belongs to class k (or 0 otherwise). The segment size ρ_k, the parameters ξ_i, B_k, Γ_k and Ψ_k of the conditional probability function are stated (as results of the M-step), and provisional estimates (expected values), $E(z_{ik}) = P_{ik}$, for z_{ik} are computed in the E-step according to Bayes' theorem (Barnard, 1958, ;E-step in Figure 2).

Table 1 Explanation of Symbols

FIMIX-PLS	
A_m	number of exogenous variables as regressors in regression m
a_m	exogenous variable a_m with $a_m = 1, \ldots, A_m$
B_m	number of endogenous variables as regressors in regression m
b_m	endogenous variable b_m with $b_m = 1, \ldots, B_m$
$\gamma_{a_m mk}$	regression coefficient of a_m in regression m for latent class k
$\beta_{b_m mk}$	regression coefficient of b_m in regression m for latent class k
τ_{mk}	$((\gamma_{a_m mk}), (\beta_{b_m mk}))'$ vector of the regression coefficients
ω_{mk}	cell $(m \times m)$ of Ψ_k
c	constant factor
$f_{i\|k}(\cdot)$	probability for case i given a latent class k and parameters $(\cdot)$
I	number of cases or observations
i	case or observation i with $i = 1, \ldots, I$
J	number of exogenous latent variables
j	exogenous latent variable j with $j = 1, \ldots, J$
K	number of latent classes
k	latent class or segment k with $k = 1, \ldots, K$
M	number of endogenous latent variables
m	endogenous latent variable m with $m = 1, \ldots, M$
N_k	number of free parameters defined as $(K-1) + KR + KM$
P_{ik}	probability of membership of case i to latent class k
R	number of predictor variables of all regressions in the inner model
S	stop or convergence criterion
V	large negative number
X_{mi}	case values of the regressors for regression m of individual i
Y_{mi}	case values of the regressant for regression m of individual i
z_{ik}	$z_{ik} = 1$, if the case i belongs to latent class k; $z_{ik} = 0$ otherwise
ζ_i	random vector of residuals in the inner model for case i
η_i	vector of endogenous latent variables in the inner model for case i
ξ_i	vector of exogenous latent variables in the inner model for case i
B	$M \times M$ path coefficient matrix of the inner model
Γ	$M \times J$ path coefficient matrix of the inner model
Δ	difference of $current_{lnL}$ and $last_{lnL}$
B_k	$M \times M$ path coefficient matrix of the inner model for latent class k
Γ_k	$M \times J$ path coefficient matrix of the inner model for latent class k
Ψ_k	$M \times M$ matrix for latent class k containing the regression variances
ρ	$(\rho_1, \ldots, \rho_K)$, vector of the K mixing proportions of the finite mixture
ρ_k	mixing proportion of latent class k
PLS-TPM	
ξ_{m*}	target endogenous latent variable
P_{m*}	number of manifest variables in the $m*$-th target block
$v^2_{ip_{m*}k}$	residual of the redundancy model for the i-th observation in the k-th latent class, corresponding to the $m*$-th target block
$Rd\,(\xi_{m*}, y_{p_{m*}})$	redundancy index for the target endogenous manifest variables for group k
REBUS	
Q	total number of latent variables, endogenous and exogenous ones with $Q = J + M$
ξ_q	generic latent variable
ξ_j	generic exogenous latent variable
ξ_m	generic endogenous latent variable
P	total number of manifest variables
P_q	number of manifest variables in the q-th block, with $\sum_{q=1}^{Q} P_q = P$
x_{pq}	generic manifest variable in the q-th block
e_{ipqk}	measurement residual for the i-th observation in the k-th latent class, corresponding to the p-th manifest variable in the q-th block, i.e. the communality residuals
f_{imk}	the structural residual for the i-th observation in the k-th latent class, corresponding to the m-th endogenous block
t_k	number of extracted components

Fig. 2 The FIMIX-PLS Algorithm

// initial E-step
set random starting values for P_{ik}; set $last_{lnL} = V$; set $0 < S < 1$

repeat do
begin
// the M-step starts here
$\rho_k = \frac{\sum_{i=1}^{I} P_{ik}}{I}, \forall k$
determine B_k, Γ_k, Ψ_k, $\forall k$
compute $current_{lnL} = \sum_i \sum_k z_{ik} ln(f(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k)) + \sum_i \sum_k z_{ik} ln(\rho_k)$
$\Delta = current_{lnL} - last_{lnL}$

// the E-step starts here
if $\Delta \geq S$ **then**
begin
$P_{ik} = \frac{\rho_k f_{i|k}(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k)}{\sum_{k=1}^{K} \rho_k f_{i|k}(\eta_i|\xi_i, B_k, \Gamma_k, \Psi_k)}, \forall i, k$
$last_{lnL} = current_{lnL}$
end
end
until $\Delta < S$

Equation 4 is maximized in the M-step. This part of the FIMIX-PLS algorithm accounts for the most important changes to fit the finite mixture approach to PLS path modeling compared with the original finite mixture structural equation modeling technique Jedidi et al. (1997). Initially, new mixing proportions ρ_k are calculated by the average of adjusted expected values P_{ik} that result from the previous E-step. Thereafter, optimal parameters for B_k, Γ_k and Ψ_k are determined by independent OLS regressions (one for each relationship between latent variables in the structural model). ML estimators of coefficients and variances are assumed to be identical to OLS predictions. The following equations are applied to obtain the regression parameters for latent endogenous variables:

$$Y_{mi} = \eta_{mi} \tag{5}$$

$$X_{mi} = (E_{mi}, N_{mi})' \tag{6}$$

$$E_{mi} = \begin{cases} \{\xi_1, \ldots, \xi_{A_m}\}, A_m \geq 1, a_m = 1, \ldots, A_m \wedge \xi_{a_m} \text{ is regressor of } m \\ \emptyset \text{ else} \end{cases} \tag{7}$$

$$N_{mi} = \begin{cases} \{\eta_1, \ldots, \eta_{B_m}\}, B_m \geq 1, b_m = 1, \ldots, B_m \wedge \eta_{b_m} \text{ is regressor of } m \\ \emptyset \text{ else} \end{cases} \tag{8}$$

The closed form OLS analytic formula for τ_{mk} and ω_{mk} is expressed as follows:

$$\tau_{mk} = ((\gamma_{a_m mk}), (\beta_{b_m mk}))' = \left[\sum_i P_{ik}(X'_{mi} X_{mi})\right]^{-1} \left[\sum_i P_{ik}(X'_{mi} Y_{mi})\right] \tag{9}$$

$$\omega_{mk} = \text{ cell } (m \times m) \text{ of } \Psi_k = \frac{\sum_i P_{ik}(Y_{mi} - X_{mi}\tau_{mk})(Y_{mi} - X_{mi}\tau_{mk})'}{I\rho_k} \tag{10}$$

As a result, the M-step determines new mixing proportions ρ_k, and the independent OLS regressions are used in the next E-step iteration to improve the outcomes for P_{ik}. The EM-algorithm stops whenever lnL hardly improves, and an a priori specified convergence criterion is reached.

The most important FIMIX-PLS computational results are the probability P_{ik}, the mixing proportions ρ_k, class-specific estimates B_k and Γ_k for the inner relationships of the path model and Ψ_k for the regression variances. In particular, with regard to the finite mixture's probabilities P_{ik} of observations to fit into the pre-determined number classes, it must be decided if FIMIX-PLS allows to detect and treat heterogeneity among consumers in the inner PLS path model estimates by (unobservable) discrete moderating factors. This is analyzed for different numbers of K classes in the next stage of the FIMIX-PLS approach.

2.1.2 Ex Post Analysis

The number of segments is usually unknown and the process of identifying an appropriate number of classes is not clear-cut when applying FIMIX-PLS. A statistically satisfactory solution for this analytical procedure does not exist for several reasons (Wedel and Kamakura, 2000). One reason is that the mixture models are not asymptotically distributed as χ^2 and disallow the likelihood ratio statistic. Consequently, the FIMIX-PLS procedure must be repeatedly performed with consecutive numbers of latent classes K (e.g. 2 to 10). Another reason is that the algorithm (Figure 2) converges for any given number of K classes—the methods "forces" the observations to fit into the given number of latent classes. Accordingly, statistically non-interpretable outcomes for the class-specific estimates B_k and Γ_k of the inner path model relationships and for Ψ_k of the regression variances for latent endogenous variables are computed when the number of classes reaches a certain level. The development of the segment sizes is a useful indicator for stopping the analysis of additional numbers of latent classes for the sake of avoiding unreasonable FIMIX-PLS results: At some point, an additional class has only a small segment size, which explains a marginal portion of heterogeneity in the overall set of data.

The emerging statistically comprehensible FIMIX-PLS estimates for different K numbers of classes are then compared for criteria such as the lnL_K, the Akaike Information Criterion (AIC_K), the AIC Controlled ($AICC_K$) or the Bayesian Information Criterion (BIC_K). These heuristic measures permit an evaluation of FIMIX-PLS computations and the quality of their segmentation. The main goal of this analysis is to capture the heterogeneity of the inner PLS path model grouping data in accordance with the FIMIX-PLS results. Within this context, the normed entropy statistic (Ramaswamy et al., 1993) is a critical criterion for analyzing class specific FIMIX-PLS results. This criterion indicates the degree of separation for all observations and their estimated membership probabilities P_{ik} on a case-by-case basis, and it subsequently reveals the most appropriate number of latent classes for segmentation:

$$EN_K = 1 - \frac{[\sum_i \sum_k -P_{ik} ln(P_{ik})]}{I ln(K)} \tag{11}$$

The EN_K is limited between 0 and 1, and the quality of separation of derived classes commensurate with the increase of this criterion. Application of FIMIX-PLS furnishes evidence that values of EN_K above 0.5 result in estimates for P_{ik} that permit unambiguous segmentation. In this case, most observations are associated with high probabilities of membership in a certain class. Hence, the segmentation always exhibits certain fuzziness and the entropy criterion is especially relevant for assessing whether a FIMIX-PLS solution is interpretable or not. In situations where a certain number of classes is identified as most appropriate—based on the heuristic evaluation—but the EN_K is considerably below 0.5, the probabilities of membership may do not allow meaningful a priori segmentation for specific PLS estimations, a comprehensible interpretation of results and sound establishment of managerial implications. Under such circumstances, and in cases where the differences between the evaluation criteria for FIMIX-PLS results of different numbers of classes only slightly differ, the highest probability per observation and its distribution regarding the entire set of data needs to be analyzed. The more that observations exhibit high membership probabilities, e.g. higher than 0.8, the better they uniquely belong to a specific class and can be well separated.

A FIMIX-PLS evaluation of PLS path modeling results gives certainty that the path model estimates are not affected by unobserved heterogeneity. Otherwise, researchers obtain an appropriate number of segments as well as information about their sizes, how well they are separable and the segments specific PLS estimates for the inner model. These FIMIX-PLS analytical results are benchmarks for selecting an explanatory variable that facilitates clustering of data (manifest variables): In accordance with the FIMIX-PLS results, data is segmented by an explanatory variable and used as new input for segment-specific LVP computations with PLS.

An explanatory variable must include both similar grouping of data, as indicated by the FIMIX-PLS results, and interpretability of the distinctive clusters. The identification of such a variable is essential to exploit FIMIX-PLS findings for PLS path modeling, and it is the most challenging analytical task to accomplish. An ex-post analysis approach by Ramaswamy et al. (1993) may supports the researcher in selecting the best fitting explanatory variable. Alternatively, if this kind of variable is not available, researchers may attain segmentation by forming groups of data according to P_{ik} and characterize them by existing manifest variables. The PLS path model is then estimated for each of the finally formed segments.

As a result, the researcher further differentiates the PLS results on the aggregate data level by identifying and forming certain groups of data with distinctive inner path model estimates facilitates multigroup PLS analyses (Chin and Dibbern, 2007). Hahn et al. (2002) and—building on their research—Ringle (2006) and Ringle et al. (2007) present in depth the FIMIX-PLS methodology, initial applications as well as problematic areas and fields of future research. Moreover, they particularly describe problematical issues linked to FIMIX-PLS, c.f. convergence of the EM algorithm in local optimum solutions, the general applicability of FIMIX-PLS on inner PLS path model constellations and the use of constrained estimations, reliable procedures for identification of an explanatory variable in the ex post analysis as well as segmentation regarding both outer and inner path model estimates.

2.2 PLS Typological Path Modeling

In comparison with FIMIX-PLS, PLS-TPM (Trinchera et al., 2006; Squillacciotti, 2007) uses a different statistical approach for response based segmentation issues. Considering that PLS path modeling aims at optimizing the models' predictivity without requiring distributional assumptions, PLS-TPM has been designed for prediction oriented path model segmentation without including distributional assumptions on latent or manifest variables level. The assignment of the units to the latent classes is achieved throughout a unit-model distance. The chosen distance is obtained as an extension of the *DModY* distance used in PLS Regression (Tenenhaus and Esposito Vinzi, 2005; Bastien et al., 2005; Tenenhaus, 1998) and in PLS Typological Regression (PLS-TR; Esposito Vinzi et al., 2004). It is computed as follows:

$$D_k = \sqrt{\frac{\sum_{p_{m*}=1}^{P_{m*}} v_{ip_{m*}k}^2 \Big/ Rd\left(\xi_{m*}, y_{p_{m*}}\right)}{\sum_{i=1}^{I_k} \sum_{p_{m*}=1}^{P_{m*}} \left[v_{ip_{m*}k}^2 \Big/ Rd\left(\xi_{m*}, y_{p_{m*}}\right)\right]}} \tag{12}$$

where ξ_{m*} is the target latent variable, $v_{ip_{m*}k}^2$ is the residual of the redundancy model (i.e., the regression of the final endogenous manifest variables over the exogenous target latent variables) and $Rd\left(\xi_{m*}, y_{p_{m*}}\right)$ represents the redundancy index for the target endogenous manifest variables for latent group k. Until today, PLS-TPM has only been tested and implemented on models containing reflective latent variables, which is a frequent situation in applications. The distance D_k has been developed for this kind of outer model. Ongoing research extends the distance measure and thereby the methodology on the subject of its utilization in formative PLS path modeling.

Since the distance is computed in the endogenous manifest variables space, the obtained segmentation optimizes the prediction power of the local models in terms of the redundancy model. PLS-TPM performs an iterative algorithm to form the latent classes. The algorithm starts with a PLS path modeling analysis on all units. Based on the results of the global model, the residuals of the redundancy model are computed for each unit. Since the number of latent classes is not known a priori, a hierarchical classification is performed on these residuals. The dendrogram resulting from the ascendant hierarchical classification allows to choose the number of segments (K) corresponding to the optimal partition as well as their initial composition. Once the initial segments defined, the K local models are estimated. We thereby obtain the segment specific PLS path modeling results that allow computing the distances (Equation 12) of each unit from each local model.

Fig. 3 The PLS-TPM Algorithm

begin
step 1: estimation of the global PLS path model
    step 1.1: computation of the redundancy residuals of all units from the global model
    step 1.2: ascendant hierarchical classification on the residuals computed at step 1.1
step 2: choice of the number of classes (K) according to the dendrogramme obtained at previous step
step 3: imputation of the units to each class according to the cluster analysis results;
repeat do
begin
    step 4: estimation of the K local models (one for each class)
    step 5: computation of the distances D_k between each unit and each local model
    step 6: assignment of each unit to the class corresponding to the closest local model
end
until convergence
step 7: description of the obtained classes according to differences among the local models
end

In every iteration, the distance of each unit from all local model is evaluated, and each unit is assigned to the closest local model with respect to the distance D_k. Modifications in the class composition leads to reestimation of the K local models as well as the distances D_k and reassignment of units. The stability of the composition of the classes or of the model coefficients from one step to the other implies convergence of the algorithm. In other words, the algorithm converges when the results at the i iteration are the same (at the 2nd decimal figure) compared to those that have been generated in one of the previous iterations, or when the composition of the classes remains unchanged from one iteration to the other. The results of the final local models are then compared with respect to the inner and outer coefficients. In PLS-TPM heterogeneity is not supposed to be limited to the inner model. Differently from FIMIX-PLS, outer coefficients are not static over the iterations but are reestimated at each step. Hence, PLS-TPM provides local models that are different not only in inner path coefficients but also in the outer relationships.

2.3 Response Based Unit Segmentation in PLS path modeling

PLS Typological Path modeling provides local models that are different with respect to both the inner and the outer models. Nonetheless, the segmentation of units is obtained only according to the inner model and to the target variable's outer model. Moreover, PLS-TPM requires to identify a target latent variable among all the endogenous latent variables. For this reason, and in order to reach a segmentation of the units that takes into account the performance of both the inner and the outer models, a new response-based segmentation technique has been recently presented in the PLS path modeling framework: the REsponse-Based Units Segmentation (REBUS; Esposito Vinzi et al., 2007).

REBUS-PLS could be considered as a development of PLS-TPM aiming to overcome some of its major drawbacks: namely, the need to identify a target latent variable and the fact to provide a segmentation according to only the inner model.

The purpose of REBUS-PLS is to identify local models that are different as regard to both the inner and the outer models' estimates and that fit better than the global model. That is why the observations are classified according to a *closeness measure* (CM) defined on the Goodness of Fit index structure (GoF; Tenenhaus et al., 2004).

$$GoF = \sqrt{\frac{\sum_{q=1}^{Q}\sum_{p=1}^{P_q} Cor^2\left(x_{pq},\xi_q\right)}{\sum_{q=1}^{Q} P_q} \times \frac{\sum_{m=1}^{M} R^2\left(\xi_m,\{\xi_j\texttt{'s explaining }\xi_m\}\right)}{M}} \tag{13}$$

where Q is the total number of latent variables (endogenous and exogenous ones), M is the number of endogenous latent variables and P_q is the number of manifest variables in the q-th block.

As for the GoF index (Equation 13), also the *closeness measure* used in REBUS-PLS is formed by two different parts (Equation 14). The left-side term refers to the observations' performance in the outer model, while the right-side term refers to the inner model. The basic idea is that if the i-th observation belongs to the k-th latent class, its performance in the local model estimated for the k-th latent class will be better than its performance in all the other local models. The performance of statistical units in the outer model is evaluated by the so called measurement or communality residuals (e_{pqk}), while their performance in the inner model is assessed by taking into account the inner or structural residuals (f_{mk}). All the computed residuals are weighted according to quality indexes: the importance of residuals increases while the quality index decreases. That is why the communality index and the R^2 values are considered in the CM.

$$CM_{ik} = \sqrt{\frac{\sum_{q=1}^{Q}\sum_{p=1}^{P_q}\left[e_{ipqk}^2/Com\left(\xi_{qk},x_{pq}\right)\right]}{\frac{\sum_i^I\sum_{q=1}^{Q}\sum_{p=1}^{P_q}\left[e_{ipqk}^2/Com\left(\xi_{qk},x_{pq}\right)\right]}{(I-t_k-1)}} \times \frac{\sum_{m=1}^{M}\left[f_{imk}^2/R^2\left(\xi_m,\{\xi_j\texttt{'s explaining }\xi_m\}\right)\right]}{\frac{\sum_i^I\sum_{m=1}^{M}\left[f_{imk}^2/R^2\left(\xi_m,\{\xi_j\texttt{'s explaining }\xi_m\}\right)\right]}{(I-t_k-1)}}} \tag{14}$$

where e_{ipqk} is the measurement residual for the i-th observation in the k-th latent class, corresponding to the p-th manifest variable in the q-th block, f_{imk} is the structural residual for the i-th observation in the k-th latent class, corresponding to the m-th endogenous block, I is the total number of observations, and t_k is the number of extracted components, since all blocks are supposed to be reflective, t_k will always be equal to 1.

For each observation, one measurement residual is computed for each manifest variable and one structural residual is computed for each endogenous variable in the inner model. For a thorough description of the REBUS-PLS algorithm and the computation of the inner and the outer residuals, please refer to the original REBUS paper (Esposito Vinzi et al., 2007).

The choice of the *CM* as a criterion for assigning observations to latent classes has two major advantages. Firstly, unobserved heterogeneity can now be detected in both the measurement and the structural model. If two models show identical inner coefficients, but differ with respect to one or more outer weights in the exogenous blocks, REBUS-PLS is able to identify this source of heterogeneity differently from PLS-TPM and FIMIX-PLS. Moreover, since the *closeness measure* is defined according to the structure of the *GoF* index, the identified local models will show a higher value for both the *GoF* and the R^2 indices.

Similarly to PLS-TPM and FIMIX-PLS, the REBUS-PLS algorithm starts by performing a PLS path model analysis on the whole set of observations. Once the global model estimated, inner and outer residuals are computed. If no a priori information is available on the number of latent classes, as it is often the case, a hierarchical cluster analysis is performed on the computed residuals. The number of latent classes to take into account, as well as the initial composition of the latent classes, is obtained by looking at the dendogramme. A local model is then estimated for each identified latent class. For each observation, a closeness measure from each local model is computed according to (Equation 14). Observations are then allocated to the closest local model, i.e. to the local model for which they show the smallest *CM* value. Changes in the latent classes composition lead to a new estimation of the local models and to a new computation of *CM*s.

The algorithm stops when less than 0.5% of units change class membership from an iteration to another. The threshold value of the 0.5% is a rule of thumb obtained according to simulation and practical applications of the REBUS-PLS algorithm in order to take into account the so called boundary-observations. If users force the algorithm to go beyond this threshols, the algorithm keeps providing the same results in terms of classes composition and local models' estimates in the successive steps.

Ex-post analysis could be performed in order to identify external (concomitant) variables that characterize the latent classes identified by REBUS-PLS. It is important to underline that local models of REBUS-PLS differ with respect to both the inner and the outer estimates. Moreover, external information is used only to describe the latent classes identified by REBUS-PLS: the latent classes' composition is obtained according to inner and outer models' residuals.

3 Design of the Numerical Example and Data Simulation

A key area for identifying and forming segments in social sciences is related to the specific behavior of certain groups of persons. Although the naming of latent variables is a trivial matter for numerical examples using simulated data, this study focuses on the area of customer satisfaction (Fornell et al., 1996) as well as segmentation of markets and consumers (Wedel and DeSarbo, 2002). We thereby identify heterogeneity and treat it by segmentation

Fig. 4 The REBUS-PLS Algorithm

begin
step 1: estimation of the global PLS path model
    step 1.1: computation of the inner and outer residuals of all observations from the global model
    step 1.2: ascendant hierarchical classification on the residuals computed at step 1.1
step 2: choice of the number of classes (K) according to the dendrogramme obtained at previous step
step 3: imputation of the observations to each class according to the cluster analysis results
repeat do
begin
    step 4: estimation of the K local models (one for each class)
    step 5: computation of the *closeness measure* CM_k between each observation and each local model
    step 6: assignment of each observation to the class corresponding to the closest local model
end
until convergence
step 7: description of the obtained classes according to differences among the local models
end

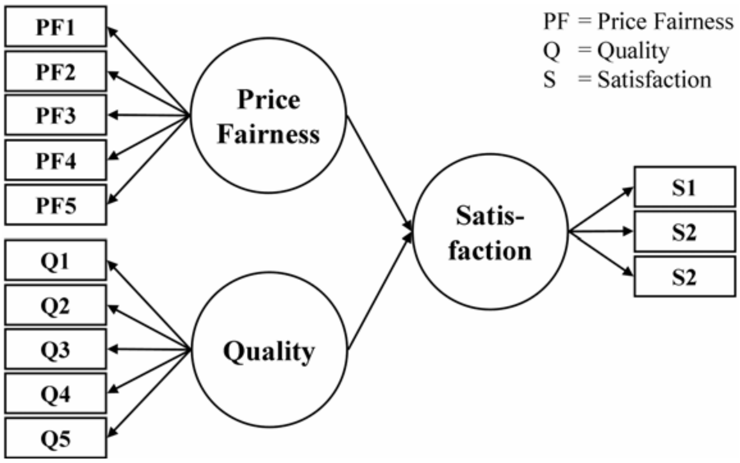
as means of presenting the general effectiveness of FIMIX-PLS and PLS-TPM for path modeling. This kind of example allows us to better illustrate the motivation and consequences of this study that are for the most part transferable to other research disciplines in social sciences.

Customer satisfaction has become a fundamental and well documented construct in business research. It is critical to demand and to any corporation's success given its importance and established relation to customer retention and corporate profitability (Anderson et al., 1994). Although it is often acknowledged that truly homogeneous segments of consumers do not exist, studies even report unobserved customer heterogeneity within a given product or service class (Wu and Desarbo, 2005). Forming groups of consumers that are homogeneous in terms of the benefits they seek or their response to marketing programs (e.g. product offering, price discounts) is therefore a key element for marketers to establish and improve their targeted marketing strategies (Wedel and Kamakura, 2000).

In terms of heterogeneity in the inner path model, it might be desirable to identify and describe price sensitive consumers or those requiring price fairness (Xia et al., 2004) and consumers who have the strongest preference for another particular product attribute (Allenby et al., 1998), e.g. quality. Thus, the path model for our numerical examples (Figure 5) has one latent endogenous variable, *Customer Satisfaction*, and two latent exogenous variables, *Price Fairness* and *Quality*, in the inner model (Jedidi et al., 1997). The experimental sets of data consist of two segments with the following characteristics: (a) Segment 1 (price fairness seeking customers) is characterized by a strong relationship between *Price Fairness* and *Customer Satisfaction* and a weak relationship between *Quality* and *Customer Satisfaction*; (b) Segment 2 (quality oriented customers) is characterized by a strong relationship between *Quality* and *Customer Satisfaction* and a weak relationship between *Price Fairness* and *Customer Satisfaction*.

Each latent exogenous variable (*Price Fairness* and *Quality*) has five manifest variables (reflective mode), and the latent endogenous variable (*Customer Satisfaction*) is measured by three indicators (reflective mode). However, it is not relevant for this study to include an additional level of complexity by exemplifying path model details regarding the manifest variables and the theoretical reasoning for choosing reflective instead of formative measurement models (c.f. Bagozzi and Edwards, 2005). PLS-TPM has been established for path models with reflective blocks and, thus, our analysis is limited to that kind of measurement model.

Fig. 5 PLS Path Model for Data Simulation



This study intentionally uses a clear cut example of a marketing related path model for data simulation purposes. We principally follow the data generation procedure that Chin et al. (2003) use for PLS path modeling. Data for each segment is first generated for the latent variables according to the relationships specified in inner path model and then data is generated for the observed variables from the latent variables in the model. This approach allows to generate data with distributional characteristics imposed by the model (Chin et al., 2003) and is consistent with the functionalities available in the SEPATH module of the software application STATISTICA 7.1 (StatSoft, Inc., 2005). Data simulation for the group of price fairness seeking consumers involves a strong relationship of 0.9 between *Price Fairness* and *Customer Satisfaction* and a week relationship of 0.1 between *Quality* and *Customer Satisfaction* in the inner path model (Segment 1). Another group of data reflects the characteristics of the quality oriented consumers (Segment 2). To start with, we simulate normal data of 100 cases per segment. Thus, data on the aggregate level for the numerical examples includes 200 cases.

An evaluation of FIMIX-PLS and PLS-TPM capabilities requires alterations of segment sizes and data distribution. Keeping the total number of cases at 200, we also simulate data for different segment proprtions, namely 0.6 to 0.4, 0.7 to 0.3 and 0.8 to 0.2. Thereby, this study exhibits whether the model based segmentation techniques for PLS correctly identify the groups when their sizes are systematically changed. Moreover, this analysis includes data that is nonnormal for the three cases of (a) skewness, (b) kurtosis and (c) both, skewness and kurtosis. Skewness has a value 0.85 and kurtosis has a value of 2.0. In accordance with other studies (c.f. Boomsma and Hoogland, 2001), the selected value of skewness and kurtosis are at moderate levels. Vale and Maurelli (1983) extended the Fleishman approach to generate multivariate nonnormal random numbers. This procedure, as it is implemented in the STATISTICA software application, fits our nonnormal data generation purposes especially since control of specified nonnormal distributions in latent variables is not of interest to us (Reinartz et al., 2002).

In total, the analysis involves 16 marketing related numerical examples on different sets of simulated data. Each set includes computation of PLS path modeling results for (a) the aggregate data level (global model) and (b) each group of simulated data (group models) as well as the class or local model solutions for (c) FIMIX-PLS employing the SmartPLS 2.0 (Ringle et al., 2005) software application and (d) PLS-TPM using a SAS macro implementation (Trinchera et al., 2006). The results of the global model exhibit the requirement for addressing heterogeneity of inner path model estimates. A comparison of the outcomes in the group and local models estimates facilitates an assessment of FIMIX-PLS and PLS-TPM. The group model estimates for the two high and low inner path model coefficients as well as R^2 of the latent endogenous variables satisfaction are benchmarks for both model based segmentation techniques. The rate of correctly assigned cases is another performance indicator. In order to evaluate that rate for FIMIX-PLS, data is classified by their probabilities of membership into the two final groups. The *EN* (Hahn et al., 2002) and the *GoF* (Tenenhaus et al., 2005) are specific evaluation criteria for each methodology that we include in our analysis. Evaluation of class separation in FIMIX-PLS uses the *EN*. This criterion requires probabilities of membership P_{ik} and, thus, is inappropriate for PLS-TPM. On the other hand, the *GoF* is a global evaluation criterion for model quality, which takes both the redundancy and the communality model into account. FIMIX-PLS cannot estimate class specific out relationships that are required for computing the *GoF*.

The design of this study allows us to present three kinds of analyses. First, we reveal the requirement and the potentials of FIMIX-PLS and PLS-TPM to detect (unobserved) heterogeneity in the inner path model relationships for different sets of simulated data. Second, the numerical examples give evidence how both methodologies perform in situations of unbalanced group sizes. Third, this study exhibits the capabilities of FIMIX-PLS and PLS-TPM to uncover the segments when data is nonnormal.

4 Simulation Study on Partial Least Squares Path Model Based Identification of Heterogeneity and Segmentation

4.1 Results on the Aggregate Data Level

Before we analyze the results of model based segmentation for FIMIX-PLS and, thereafter, for PLS-TPM, it is important to illustrate the requirement for these kinds of analyses. The PLS path modeling results on the aggregate data level (Figure 6 to Figure 9) are significantly different compared with the segment specific computations for each a priori simulated group of data. In these numerical examples, estimates for the overall set of data are close to the weighted average of group specific coefficients. As a consequence, the PLS path modeling results are ambiguous when heterogeneity is not accounted for, especially when distinctive groups have about the same size. For instance, in the numerical example for the 0.5 to 0.5 segment sizes case, the two inner relationships of about 0.9 and 0.1 for one group and vice versa for the other group of simulated data turn out to have a value on the aggregate data level between 0.450 and 0.550 for both relationships. Moreover, R^2 is significantly lower than for the PLS estimations for each group of data.

If heterogeneity is not identified by the researcher, the effects of *Price Fairness* and *Quality* on *Satisfaction* seem to be equally important, at least in the numerical examples with equal segments sizes. As a consequence of these PLS path modeling results, marketers may focus on the areas of *Price Fairness* and *Quality* at the same time for all consumers. Uncovering heterogeneity the inner path model relationships and forming distinctive groups of price fairness seeking and quality oriented customers allows marketers to develop better targeted and more effective business strategies. However, the requirement for model based

Fig. 6 FIMIX-PLS and PLS-TPM Results for Simulated Data (Part A)

Data Generation Scheme for Manifest Variables		Segment Size 0.5	0.5	Segment Size 0.5	0.5	Segment Size 0.6	0.4	Segment Size 0.6	0.4
		Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2
Method		FIMIX-PLS		PLS-TPM		FIMIX-PLS		PLS-TPM	
Normal Distribution	PF --> Sat.	0.899	0.113	0.916	0.019	0.918	0.129	0.898	0.274
		(0.873*)	(0.169*)	(0.873*)	(0.169*)	(0.879*)	(0.201*)	(0.879*)	(0.201*)
		(0.538**)		(0.538**)		(0.620**)		(0.620**)	
	Q --> Sat.	0.009	0.908	0.052	0.901	0.006	0.859	0.041	0.517
		(0.078*)	(0.898*)	(0.078*)	(0.898*)	(0.076*)	(0.896*)	(0.076*)	(0.896*)
		(0.450**)		(0.450**)		(0.374**)		(0.374**)	
	R² (Sat.)	0.831	0.851	0.846	0.838	0.836	0.857	0.782	0.371
		(0.777*)	(0.843*)	(0.777*)	(0.843*)	(0.789*)	(0.851*)	(0.789*)	(0.851*)
		(0.465**)		(0.465**)		(0.511**)		(0.511**)	
	GoF	-	-	0.425	0.360	-	-	0.460	0.233
		(0.734*)	(0.776*)	(0.734*)	(0.776*)	(0.759*)	(0.692*)	(0.759*)	(0.692*)
		(0.593**)		(0.593**)		(0.679**)		(0.679**)	
	EN	0.540		-		0.559		-	
	Segment Size	0.519	0.481	0.555	0.445	0.586	0.414	0.530	0.470
	Correctly Assigned Cases	0.850	0.780	0.712	0.764	0.750	0.757	0.859	0.266
		0.815		0.735		0.754		0.580	
Non-Normal Data (Skewness)	PF --> Sat.	0.847	0.031	0.926	0.022	0.931	0.109	0.937	0.096
		(0.886*)	(0.080*)	(0.886*)	(0.080*)	(0.879*)	(0.033*)	(0.879*)	(0.033*)
		(0.511**)		(0.511**)		(0.567**)		(0.567**)	
	Q --> Sat.	0.091	0.881	0.105	0.834	0.070	0.826	0.035	0.821
		(0.125*)	(0.878*)	(0.125*)	(0.878*)	(0.139*)	(0.893*)	(0.139*)	(0.893*)
		(0.468**)		(0.468**)		(0.417**)		(0.417**)	
	R² (Sat.)	0.725	0.853	0.755	0.860	0.835	0.751	0.883	0.724
		(0.786*)	(0.782*)	(0.786*)	(0.782*)	(0.786*)	(0.788*)	(0.786*)	(0.788*)
		(0.442**)		(0.442**)		(0.470**)		(0.470**)	
	GoF	-	-	0.344	0.461	-	-	0.499	0.319
		(0.773*)	(0.706*)	(0.773*)	(0.706*)	(0.782*)	(0.692*)	(0.782*)	(0.692*)
		(0.579**)		(0.579**)		(0.623**)		(0.623**)	
	EN	0.467		-		0.440		-	
	Segment Size	0.541	0.459	0.465	0.585	0.525	0.475	0.550	0.450
	Correctly Assigned Cases	0.770	0.770	0.699	0.729	0.733	0.757	0.718	0.767
		0.770		0.715		0.742		0.740	

* PLS computational results for each group of simulated data
** PLS computational results for the aggregate data level
PF = Price Fairness / Q = Quality / Sat. = Satisfaction

segmentation decreases when on segment dominates the other and thereby the path model estimates on the aggregate data level.

4.2 Result for the Finite Mixture Partial Least Squares Approach

The simulation study reveals that the FIMIX-PLS methodology is capable to reliably uncover heterogeneity in all 16 examples under evaluation (Figure 6 to Figure 9). This finding is based on the FIMIX-PLS results for mixing proportions, segments specific path coefficients, and the R^2 of *Satisfaction*. Most important, the FIMIX-PLS mixing proportions, which are computed including the probabilities of membership of each case (Figure 2), and inner path model coefficients for each class are very close to the pre-determined segment sizes and relationships in all numerical examples. Moreover, with respect to the assumption that FIMIX-PLS aims at separating segment specific distributional functions with conditional multivariate normal densities (Hahn et al., 2002), it is essential to note that the identification of segments is very robust regarding violations of the multivariate distributional assumption. FIMIX-PLS appropriately uncovers the two a priori formed segments even in situations when data is extremely nonnormal (the case of skewness and kurtosis).

Fig. 7 FIMIX-PLS and PLS-TPM Results for Simulated Data (Part B)

Data Generation Scheme for Manifest Variables		Segment Size 0.5	0.5	Segment Size 0.5	0.5	Segment Size 0.6	0.4	Segment Size 0.6	0.4
		Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2
Method		**FIMIX-PLS**		**PLS-TPM**		**FIMIX-PLS**		**PLS-TPM**	
Non-Normal Data (Kurtosis)	PF --> Sat.	0.887 (0.884*)	0.074 (0.062*)	0.876 (0.884*)	0.048 (0.062*)	0.898 (0.885*)	0.095 (0.087*)	0.914 (0.885*)	0.038 (0.087*)
		(0.506**)		(0.506**)		(0.613**)		(0.613**)	
	Q --> Sat.	0.157 (0.139*)	0.880 (0.896*)	0.076 (0.139*)	0.901 (0.896*)	0.141 (0.118*)	0.812 (0.863*)	0.057 (0.118*)	0.828 (0.863*)
		(0.482**)		(0.482**)		(0.360**)		(0.360**)	
	R^2 (Sat.)	0.769 (0.818*)	0.817 (0.808*)	0.782 (0.818*)	0.821 (0.808*)	0.759 (0.815*)	0.785 (0.763*)	0.840 (0.815*)	0.698 (0.763*)
		(0.513**)		(0.513**)		(0.544**)		(0.544**)	
	GoF	- (0.807*)	- (0.779*)	0.473 (0.807*)	0.395 (0.779*)	- (0.805*)	- (0.756*)	0.453 (0.805*)	0.402 (0.756*)
		(0.628**)		(0.628**)		(0.731**)		(0.731**)	
	EN	0.409		-		0.391		-	
	Segment Size	0.525	0.475	0.520	0.480	0.623	0.377	0.440	0.560
	Correctly Assigned Cases	0.810	0.790	0.750	0.771	0.792	0.643	0.841	0.589
		0.800		0.760		0.737		0.700	
Non-Normal Data (Skewness and Kurtosis)	PF --> Sat.	0.924 (0.894*)	0.069 (0.057*)	0.907 (0.894*)	0.054 (0.057*)	0.929 (0.890*)	0.060 (0.091*)	0.902 (0.890*)	0.035 (0.091*)
		(0.489**)		(0.489**)		(0.544**)		(0.544**)	
	Q --> Sat.	0.097 (0.052*)	0.784 (0.910*)	0.009 (0.052*)	0.907 (0.910*)	0.087 (0.064*)	0.784 (0.927*)	0.047 (0.064*)	0.645 (0.927*)
		(0.541**)		(0.541**)		(0.474**)		(0.474**)	
	R^2 (Sat.)	0.779 (0.809*)	0.851 (0.814*)	0.825 (0.809*)	0.851 (0.814*)	0.786 (0.808*)	0.876 (0.830*)	0.844 (0.808*)	0.403 (0.830*)
		(0.516**)		(0.516**)		(0.509**)		(0.509**)	
	GoF	- (0.769*)	- (0.793*)	0.460 (0.769*)	0.401 (0.793*)	- (0.768*)	- (0.797*)	0.399 (0.768*)	0.316 (0.797*)
		(0.625**)		(0.625**)		(0.668**)		(0.668**)	
	EN	0.394		-		0.422		-	
	Segment Size	0.497	0.503	0.535	0.465	0.591	0.409	0.420	0.580
	Correctly Assigned Cases	0.730	0.800	0.748	0.785	0.767	0.657	0.750	0.509
		0.765		0.765		0.726		0.610	

* PLS computational results for each group of simulated data
** PLS computational results for the aggregate data level
PF = Price Fairness / Q = Quality / Sat. = Satisfaction

This study on simulated data does not include an explanatory variable for final segmentation according to stage two of a comprehensive FIMIX-PLS analysis (Figure 1). As an alternative, we use the P_{ik} results to assign each case to one of the two segments and, thereby, to assess their correct assignment in each numerical example. The overall rate of correctly assigned cases ranges from 0.726 to 0.905. This is a sufficiently high level to positively evaluate the capabilities of FIMIX-PLS considering that data simulation includes a certain number of cases that fit well in one and in the other group and represent noisy data in the intersection of both generated segments. The *EN* computation includes the probabilities of membership to indicate how well classes are separable. A value of EN above 0.5 in the different sets of normal data with altered segments sizes gives evidence that the observations have relatively high P_{ik} and, thus, the data is well separable into two groups (Ringle et al., 2007). The decline of EN below 0.5 in numerical examples for nonnormal data reveals that classification becomes more ambiguous in terms of P_{ik}. A comparison of the three cases of nonnormal data discloses that ”skewness” provides better results than ”kurtosis” regarding the rate of correctly assigned cases while ”kurtosis” performs better than ”skewness” in terms of EN. The combination of kurtosis and skewness joins both weak effects.

Regardless whether data is normal or nonnormal, in the numerical examples for the 0.5 to 0.5 segment sizes, FIMIX-PLS provides class specific path coefficients that are comparable to those PLS estimates for each a priori formed segment and an in average increased R^2. However, the methodology has a tendency to produce more distinctive segment specific path coefficients and a higher R^2 outcome for one segment when the segment sizes are systematically changed. This tendency towards capitalizing on certain segments is determined by the FIMIX-PLS methodology. In particular, the most important determinants are (a) the segment sizes as well as, (b) the heterogeneity and, thus, the distinctiveness of data for separation regarding the path coefficients, and (c) the level of R^2 in the PLS estimates for each group of simulated data. The more those criteria direct the algorithm towards a particular segment, the higher are its FIMIX-PLS differences of segment specific path coefficients and the increase of R^2 compared to the results for each a priori formed segment.

As a consequence, the rate of correctly assigned cases increases for the larger segment while that for the smaller segments declines resulting in an enhanced overall appropriateness of assignment in this study. We finally find that in the examples of normal and kurtic data P_{ik} is relatively high when the cases are correctly clustered. Otherwise, P_{ik} is close to 0.5 in these numerical examples for two classes. Accordingly, EN systematically rises when the segment sizes further diverge considering the methodology’s tendency to increase the percentage of correctly assigned cases for the larger segment. However, this kind of outcome does not hold when skewness is involved. FIMIX-PLS computations become more fuzzy regarding P_{ik} and, thus, EN is at lower levels. Moreover, increases (decreases) of the size of the smaller (larger) segment results in a cutback of the effect that the cases of the larger segment improve in their rate of correct assignment at cost of the smaller segment.

These finding support the reliability and robustness of FIMIX-PLS. In situations when nonnormality of data becomes an issue, the researcher must carefully consider the effects of kurtosis and skewness on the FIMIX-PLS results. The reliability of results depends on the degree of inner path model heterogeneity as well as the nonnormality of data. Our study gives evidence that FIMIX-PLS clearly indicates potential problems of unobserved heterogeneity even in extreme data constellation.

4.3 Segmentation Using Partial Least Squares Typological Path Modeling

In all data sets PLS-TPM unfolds two classes: one shows a high structural relationship from *Price Fairness* to *Satisfaction* and a non significant relationship between *Quality* and *Satisfaction* while the other one shows exactly the opposite characteristics. The original simulation scheme is thus respected. All local models lead to a better model performance in terms of R^2 compared to the global model. However, the final local models results show globally lower values for the GoF. This is due to the chosen distance measure, which optimizes the local redundancy models’ predictivity (R^2) and does not take the exogenous variables’ local

Fig. 8 FIMIX-PLS and PLS-TPM Results for Simulated Data (Part C)

Data Generation Scheme for Manifest Variables		Segment Size 0.7	0.3	Segment Size 0.7	0.3	Segment Size 0.8	0.2	Segment Size 0.8	0.2
		Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2
Method		**FIMIX-PLS**		**PLS-TPM**		**FIMIX-PLS**		**PLS-TPM**	
Normal Distribution	PF --> Sat.	0.876	0.012	0.931	0.194	0.923	0.078	0.938	0.126
		(0.877*)	(0.146*)	(0.877*)	(0.146*)	(0.879*)	(0.296*)	(0.879*)	(0.296*)
		(0.678**)		(0.678**)		(0.759**)		(0.759**)	
	Q --> Sat.	0.041	0.906	0.070	0.567	0.036	0.814	0.060	0.413
		(0.113*)	(0.883*)	(0.113*)	(0.883*)	(0.076*)	(0.895*)	(0.076*)	(0.895*)
		(0.296**)		(0.296**)		(0.215**)		(0.215**)	
	R² (Sat.)	0.821	0.868	0.872	0.406	0.824	0.869	0.900	0.210
		(0.783*)	(0.823*)	(0.783*)	(0.823*)	(0.789*)	(0.835*)	(0.789*)	(0.835*)
		(0.531**)		(0.531**)		(0.599**)		(0.599**)	
	GoF	-	-	0.487	0.246	-	-	0.460	0.163
		(0.750*)	(0.775*)	(0.750*)	(0.775*)	(0.652*)	(0.782*)	(0.652*)	(0.782*)
		(0.634**)		(0.634**)		(0.682**)		(0.682**)	
	EN	0.619		-		0.641		-	
	Segment Size	0.707	0.293	0.510	0.490	0.769	0.231	0.490	0.510
	Correctly Assigned Cases	0.900	0.567	0.794	0.398	0.913	0.600	0.859	0.302
		0.800		0.600		0.850		0.580	
Non-Normal Data (Skewness)	PF --> Sat.	0.919	0.043	0.955	0.132	0.902	0.242	0.947	0.098
		(0.884*)	(0.068*)	(0.884*)	(0.068*)	(0.875*)	(0.245*)	(0.875*)	(0.245*)
		(0.647**)		(0.647**)		(0.751**)		(0.751**)	
	Q --> Sat.	0.035	0.777	0.007	0.691	0.097	0.693	0.088	0.553
		(0.093*)	(0.909*)	(0.093*)	(0.909*)	(0.137*)	(0.864*)	(0.137*)	(0.864*)
		(0.319**)		(0.319**)		(0.284**)		(0.284**)	
	R² (Sat.)	0.845	0.731	0.913	0.564	0.831	0.602	0.903	0.344
		(0.799*)	(0.801*)	(0.799*)	(0.801*)	(0.768*)	(0.856*)	(0.768*)	(0.856*)
		(0.477**)		(0.477**)		(0.597**)		(0.597**)	
	GoF	-	-	0.487	0.277	-	-	0.534	0.189
		(0.720*)	(0.777*)	(0.720*)	(0.777*)	(0.752*)	(0.723*)	(0.752*)	(0.723*)
		(0.602**)		(0.602**)		(0.675**)		(0.675**)	
	EN	0.497		-		0.430		-	
	Segment Size	0.627	0.373	0.520	0.480	0.726	0.274	0.560	0.440
	Correctly Assigned Cases	0.850	0.700	0.865	0.479	0.931	0.475	0.893	0.318
		0.805		0.680		0.840		0.640	

* PLS computational results for each group of simulated data
** PLS computational results for the aggregate data level
PF = Price Fairness / Q = Quality / Sat. = Satisfaction

models into account. A generalization of the distance measure in order to also account for exogenous blocks' communality is subject of ongoing research.

In data sets where groups are mixed in equal proportions, well classified rates are in all cases higher than 0.72. The worst results are obtained for the data generated with skewness, and the best result for the data characterized by kurtosis. In the data sets with groups showing different mixing proportions, the well classified rates for all data sets become on the whole lower, especially when the size of one group becomes very small compared to the other. In such situations, one of the two groups has a far higher weight in the definition of the structural coefficients of the global model instead of having a starting sample where quality oriented and price fairness seeking customers are mixed in equal proportions and, hence, a global model resulting as an "average" of the two groups. In other words, the global model is already price fairness seeking, making it harder to "isolate" the units with a different behavior. This finding is substantiated by the results for constellations where well classified rates are generally higher in the largest group.

Fig. 9 FIMIX-PLS and PLS-TPM Results for Simulated Data (Part D)

Data Generation Scheme for Manifest Variables		Segment Size 0.7	0.3	Segment Size 0.7	0.3	Segment Size 0.8	0.2	Segment Size 0.8	0.2
		Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2	Segment 1	Segment 2
Method		**FIMIX-PLS**		**PLS-TPM**		**FIMIX-PLS**		**PLS-TPM**	
Non-Normal Data (Kurtosis)	PF --> Sat.	0.874 (0.881*)	0.047 (0.059*)	0.924 (0.881*)	0.072 (0.059*)	0.879 (0.882*)	0.058 (0.067*)	0.922 (0.882*)	0.019 (0.067*)
		(0.646**)		(0.646**)		(0.713**)		(0.713**)	
	Q --> Sat.	0.160 (0.132*)	0.877 (0.917*)	0.050 (0.132*)	0.842 (0.917*)	0.155 (0.140*)	0.922 (0.940*)	0.106 (0.140*)	0.792 (0.940*)
		(0.361**)		(0.361**)		(0.296**)		(0.296**)	
	R^2 (Sat.)	0.771 (0.810*)	0.890 (0.853*)	0.857 (0.810*)	0.735 (0.853*)	0.787 (0.820*)	0.906 (0.890*)	0.864 (0.820*)	0.635 (0.890*)
		(0.575**)		(0.575**)		(0.628**)		(0.628**)	
	GoF	- (0.797*)	- (0.785*)	0.461 (0.797*)	0.420 (0.785*)	- (0.809*)	- (0.829*)	0.548 (0.809*)	0.400 (0.829*)
		(0.666**)		(0.666**)		(0.702**)		(0.702**)	
	EN	0.559		-		0.643		-	
	Segment Size	0.747	0.253	0.450	0.550	0.814	0.186	0.485	0.515
	Correctly Assigned Cases	0.971	0.533	0.944	0.500	0.994	0.550	0.959	0.313
		0.840		0.700		0.905		0.645	
Non-Normal Data (Skewness and Kurtosis)	PF --> Sat.	0.895 (0.897*)	0.386 (0.041*)	0.931 (0.897*)	0.061 (0.041*)	0.903 (0.879*)	0.331 -0.034	0.930 (0.903*)	0.127 -0.034
		(0.630**)		(0.630**)		(0.700**)		(0.700**)	
	Q --> Sat.	0.022 (0.069*)	0.717 (0.911*)	0.110 (0.069*)	0.684 (0.911*)	0.029 (0.076*)	0.647 (0.901*)	0.028 (0.031*)	0.767 (0.901*)
		(0.416**)		(0.416**)		(0.299**)		(0.299**)	
	R^2 (Sat.)	0.902 (0.822*)	0.591 (0.815*)	0.906 (0.822*)	0.486 (0.815*)	0.905 (0.789*)	0.441 (0.827*)	0.873 (0.82*)	0.616 (0.827*)
		(0.566**)		(0.566**)		(0.569**)		(0.569**)	
	GoF	- (0.781*)	- (0.802*)	0.460 (0.781*)	0.321 (0.802*)	- (0.773*)	- (0.826*)	0.489 (0.773*)	0.319 (0.826*)
		(0.654**)		(0.654**)		(0.654**)		(0.654**)	
	EN	0.330		-		0.442		-	
	Segment Size	0.555	0.445	0.455	0.545	0.693	0.307	0.520	0.480
	Correctly Assigned Cases	0.836	0.600	0.857	0.431	0.913	0.550	0.952	0.365
		0.765		0.625		0.840		0.670	

* PLS computational results for each group of simulated data
** PLS computational results for the aggregate data level
PF = Price Fairness / Q = Quality / Sat. = Satisfaction

The PLS-TPM performs very efficiently in the numerical examples under evaluation. The highest number of iterations (19) before convergence was required for data generated with high skewness. In all three remaining data sets, less than 14 iterations led to convergence.

4.4 Comparison of Results

FIMIX-PLS and PLS-TPM are the first PLS related statistical approaches for model based identification of heterogeneity. Both methodologies are able to detect the two latent classes in all numerical examples using different sets of simulated data. However, their statistical foundations are diametrical unalike. Our primary analysis and comparison reveals certain characteristics that are important for researchers and practitioners when evaluating PLS path modeling results by employing FIMIX-PLS and PLS-TPM. In general, advantages of one approach are disadvantages of the other.

FIMIX-PLS uses underlying distributional assumptions to form the latent classes. As a result, the procedure provides the probabilities of membership for each case to belong the k latent class. Computations of all other FIMIX-PLS results are based on P_{ik}. Even though exhibiting a strong robustness, nonnormality of data negatively affects the method's performance. In contrast, PLS-TPM estimates the model parameters for each segment in every iteration and units are unambiguously assigned to the closest local model regarding the distance defined in Equation 12. The researcher or practitioners does not obtain P_{ik} related results but the final separation into K distinctive (non overlapping) sets of data and their segment specific PLS path model estimates. PLS-TPM is a distribution-free method. Results are on similar levels regardless whether data is normal or nonnormal. However, PLS-TPM uses a distance measure that exhibits the tendency to form groups of equal size. Uneven segment sizes have a negative effect on the performance of this approach. In contrast, FIMIX-PLS estimations for each class and segments sizes usually meet the expectations regarding the a priori formed groups of data in the numerical examples. However, certain determinants, e.g. unbalanced segments sizes, cause a tendency to prefer one segment above the other. The FIMIX-PLS results for the larger class turn out to be slightly better in this study. This behavior also holds true in PLS-TPM.

Both FIMIX-PLS and PLS-TPM account for heterogeneity in the inner path model relationships. It is desirable that model based segmentation also identifies segment specific differences in the formative outer model relationships. This mode of PLS estimations is of high relevance in analyses, for example, on success factors in the strategic management or marketing disciplines. PLS-TPM is a key development into that direction. The methodology uses a distance measure that focuses on the redundancy model, reestimates the entire model in each iteration and, thus, considers the specific outer relationships of each local model when forming segments. But in the current state of development, the methodology is only applicable to path models involving reflective measurement models. Future research on the distance measure as well as on the optimization criterion and its behavior must exploit the potential capabilities of PLS-TPM to accomplish segmentation for all path model relationships. On the contrary, the FIMIX-PLS optimization criterion is definite and only aims at the inner model estimates. The algorithm uses static latent variable scores as input for latent class segmentation (Figure 1). As a consequence, the methodology has no capabilities to explicitly account for heterogeneity in outer relationships. This is a drawback in practical applications but allows for general applicability to PLS path modeling. FIMIX-PLS reliably identifies heterogeneity in inner path model relationships regardless whether outer measurement model operationalization is in the formative or reflective mode (Ringle et al., 2007).

5 Conclusion and Future Research

The numerical examples on customer satisfaction demonstrate that interpretation of PLS path modeling results on the aggregate data level is misleading whenever they involve heterogeneity inner model estimates. Our study gives evidence that FIMIX-PLS and PLS-TPM reliably identify this kind of heterogeneity and guide separation of the two a priori created groups of price fairness seeking and quality oriented customers. The model based segmentation techniques provide additional analytical findings where conventional clustering methods fail. Researchers and practitioners can exploit these PLS model based segmentation techniques to assure that results on the aggregate data level are not affected by unobserved heterogeneity in the inner path model estimates. Otherwise, FIMIX-PLS and PLS-TPM indicate how to treat that problem by forming groups of data. A multigroup PLS path analysis on the finally established segments (Chin and Dibbern, 2007) gives evidence if segment specific estimates are significantly different. In this case, researchers or practitioners obtain supplementary analytical results to single out their conclusions, in example, to form business strategies that effectively target certain customer segments. We believe that FIMIX-PLS and PLS-TPM are key extensions of PLS path modeling and will become a standard procedures for evaluating results.

Our initial assessment and comparison of both methodologies reveals certain strengths of each approach that are by some extent weaknesses of the other. We find that FIMIX-PLS is a generally applicable method for PLS path modeling, accurately identifies the class sizes and determines their distinctiveness in terms of EN. In contrast, we conclude that PLS-TPM does not require distributional assumptions, dynamically updates group specific PLS estimations and, thereby, provides a final assignment of observations to segments that is not based on probabilities of membership. In view of these different characteristics, a combined application of both methodologies in PLS path modeling is most effective to researchers and practitioners. In the first stage, FIMIX-PLS indicates whether heterogeneity in inner path model estimates represents a problem for the PLS analysis by rendering an appropriate number of groups, their sizes, and how well they can be separated regarding EN as well as segment specific inner model estimates. In the next stage, PLS-TPM provides segmentation and PLS estimations for each group of data. We finally conclude that the PLS-TPM methodology better fits the statistical characteristics of PLS path modeling and has more potentials for extensions to improve model based segmentation.

The present work is an initial contribution on the subject of segmentation techniques in PLS path modeling. In this sense, many aspects of the methodologies require further investigations. Future research on PLS-TPM must extend the methodology on formative measurement models and, thereby, address its general applicability to PLS path modeling. More specifically, the distance measure that determines the optimization criterion of classification performed through PLS-TPM and the convergence properties of the PLS-TPM algorithm requires further analyses. An additional evaluation criterion, similar to EN for FIMIX-PLS, must indicate the distinctiveness of distances for the assignment of cases to the final local models. Future research on FIMIX-PLS must address known problems regarding, for example, local optimum solutions or non-interpretable FIMIX-PLS estimates (Ringle et al., 2007). A critical issue of both approaches is the identification of an explanatory variable that allows the kind of segmentation as indicated by FIMIX-PLS or PLS-TPM. The development of appropriate methodologies is essential to complement model based segmentation. Moreover, a first numerical assessment and comparison of both approaches cannot state the general validity of our conclusions. Our analysis is limited to one kind of path model, two segments and a single set of data for each numerical example. Future research must present extensive simulation studies with experimental data and a broad use of empirical data to assess FIMIX-PLS and PLS-TPM capabilities for a variety of data and path model constellations.

References

Allenby, G. M., Arora, N., Ginter, J. L., 1998. On the heterogeneity of demand. Journal of Marketing Research 35 (3), 384–389.

Anderson, E. W., Fornell, C., Lehmann, D. R., 1994. Customer satisfaction, market share, and profitability: findings from Sweden. Journal of Marketing 58 (3), 53–66.

Bagozzi, R. P., Edwards, J. R., 2005. On the nature and direction of relationships between constructs and measures. Psychological Methods 2 (5), 155–174.

Barnard, G. A., 1958. Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. Biometrika 45 (3/4), 293–315.

Bastien, P., Tenenhaus, M., Esposito Vinzi, V., 2005. PLS generalised linear regression. Computational Statistics & Data Analysis 48 (1), 17–46.

Boomsma, A., Hoogland, J. J., 2001. The robustness of LISREL modeling revisited. In: Cudeck, R., du Toit, S., Sörbom, D. (Eds.), Structural Equation Modeling: Present and Future. Scientific Software International, Chicago, pp. 139–168.

Chin, W. W., 1998. Issues and opinion on structural equation modeling. MIS Quarterly 22 (1), XII–XVI.

Chin, W. W., Dibbern, J., 2007. A permutation based procedure for multi-group PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA. In: Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields - Forthcoming. Springer, Berlin et al.

Chin, W. W., Marcolin, B. L., Newsted, P. N., 2003. A partial least squares approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic mail emotion/adoption study. Information Systems Research 14 (2), 189–217.

Diamantopoulos, A., Winklhofer, H., 2001. Index construction with formative indicators: an alternative to scale development. Journal of Marketing Research 38 (2), 269–277.

Esposito Vinzi, V., Lauro, C., Amato, S., 2004. PLS typological regression: Algorithmic, classification and validation issues. In: Vichi, M., Monari, P., Mignani, S., Montanari, A. (Eds.), New Developments in Classification and Data Analysis. Springer, Berlin et al., pp. 133–140.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., Tenenhaus, M., 2007. REBUS-PLS: A response - based procedure for detecting unit segments in pls path modeling. submitted to Applied Stochastic Models in Business and Industry (ASMBI).

Fornell, C., Bookstein, F. L., 1982. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. Journal of Marketing Research 19 (4), 440–452.

Fornell, C., Johnson, M. D., Anderson, E. W., Jaesung, C., Bryant, B. E., 1996. The American customer satisfaction index: Nature, purpose, and findings. Journal of Marketing 60 (4), 7–18.

Fornell, C., Lorange, P., Roos, J., 1990. The cooperative venture formation process: A latent variable structural modeling approach. Management Science 36 (10), 1246–1255.

Fornell, C., Robinson, W. T., Wernerfelt, B., 1985. Consumption experience and sales promotion expenditure. Management Science 31 (9), 1084–1105.

Gerbing, D. W., Anderson, J. C., 1988. An updated paradigm for scale development incorporating unidimensionality and its assessment. Journal of Marketing Research 25 (2), 186–192.

Gray, P. H., Meister, D. B., 2004. Knowledge sourcing effectiveness. Management Science 50 (6), 821–834.

Hahn, C., Johnson, M. D., Herrmann, A., Huber, F., 2002. Capturing customer heterogeneity using a finite mixture PLS approach. Schmalenbach Business Review 54 (3), 243–269.

Henseler, J., Fassott, G., 2007. Testing moderating effects in PLS path models: An illustration of available procedures. In: Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields - Forthcoming. Springer, Berlin et al.

Jedidi, K., Jagpal, H. S., DeSarbo, W. S., 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. Marketing Science 16 (1), 39–59.

Jöreskog, K. G., 1978. Structural analysis of covariance and correlation matrices. Psychometrika 43 (4), 443–477.

Lohmöller, J.-B., 1989. Latent Variable Path Modeling with Partial Least Squares. Physica, Heidelberg.

Marcoulides, G. A., Saunders, C., 2006. PLS: A silver bullet? MIS Quarterly 30 (2), iii–ix.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., Robinson, W. T., 1993. An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Marketing Science 12 (1), 103–124.

Reinartz, W. J., Echambadi, R., Chin, W. W., 2002. Generating non-normal data for simulation of structural equation models using mattson's method. Multivariate Behavioral Research 37 (2), 227–244.

Rigdon, E. E., 1998. Structural equation modeling. In: Marcoulides, G. A. (Ed.), Modern Methods for Business Research. Erlbaum, Mahwah, pp. 251–294.

Ringle, C. M., 2006. Segmentation for path models and unobserved heterogeneity: The finite mixture partial least squares approach. Research Papers on Marketing and Retailing No. 35, Hamburg.

Ringle, C. M., Wende, S., Will, A., 2005. SmartPLS 2.0. http://www.smartpls.de.

Ringle, C. M., Wende, S., Will, A., 2007. Finite mixture partial least squares analysis: Methodology and numerical examples. In: Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields - Forthcoming. Springer, Berlin et al.

Squillacciotti, S., 2007. Prediction-oriented classification in PLS path modeling. In: Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (Eds.), Handbook of Partial Least Squares: Concepts, Methods and Applications in Marketing and Related Fields - Forthcoming. Springer, Berlin et al.

StatSoft, Inc., 2005. STATISTICA for Windows Version 7.1. http://www.statsoft.com.

Tenenhaus, M., 1998. La régression PLS. Technip, Paris.

Tenenhaus, M., Amato, S., Esposito Vinzi, V., 2004. A global goodness-of-fit index for pls structural equation modelling. In: Proceedings of the XLII SIS Scientific Meeting,Contributed Papers. CLEUP, Padova, pp. 739–742.

Tenenhaus, M., Esposito Vinzi, V., 2005. PLS regression, PLS path modeling and generalized Procrustean analysis: a combined approach for multiblock analysis. Journal of Chemometrics 19 (3), 145–153.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., Lauro, C., 2005. PLS path modeling. Computational Statistics & Data Analysis 48 (1), 159–205.

Trinchera, L., Squillacciotti, S., Esposito Vinzi, V., 2006. PLS typological path modeling: A model-based approach to classification. In: Esposito Vinzi, V., Lauro, C., Braverman, A., Schimek, M. G., Kiers, H. A. (Eds.), Electronic Proceedings of Knowledge Extraction and Modeling Conference.

Vale, C. D., Maurelli, V. A., 1983. Simulating multivariate nonnormal distributions. Psychometrika 48 (3), 465–471.

Venkatesh, V., Agarwal, R., 2006. Turning visitors into customers: a usability-centric perspective on purchase behavior in electronic channels. Management Science 52 (3), 367–382.

Wedel, M., DeSarbo, W. S., 2002. Market segmentation derivation and profiling via a finite mixture model framework. Marketing Letters 13 (1), 17–25.

Wedel, M., Kamakura, W., 2000. Market Segmentation: Conceptual and Methodological Foundations, 2nd Edition. Springer, New York.

Wold, H., 1982. Soft modeling: The basic design and some extensions. In: Jöreskog, K. G., Wold, H. (Eds.), Systems under Indirect Observations. North-Holland, Amsterdam, pp. 1–54.

Wu, J., Desarbo, W. S., 2005. Market segmentation for customer satisfaction studies via a new latent structure multidimensional scaling model. Applied Stochastic Models in Business and Industry 21 (4/5), 303–309.

Xia, L., Monroe, K. B., Cox, J. L., 2004. The price is unfair! A conceptual framework of price fairness perceptions. Journal of Marketing 68 (4), 1–15.

ESSEC

CENTRE
DE RECHERCHE

LISTE DES DOCUMENTS DE RECHERCHE DU CENTRE DE RECHERCHE DE L'ESSEC
(Pour se procurer ces documents, s'adresser au CENTRE DE RECHERCHE DE L'ESSEC)

LISTE OF ESSEC RESEARCH CENTER WORKING PAPERS
(Contact the ESSEC RESEARCH CENTER for information on how to obtain copies of these papers)

RESEARCH.CENTER@ESSEC.FR

2004

04001 ***BESANCENOT Damien, VRANCEANU Radu***
Excessive Liability Dollarization in a Simple Signaling Model

04002 ***ALFANDARI Laurent***
Choice Rules Size Constraints for Multiple Criteria Decision Making

04003 ***BOURGUIGNON Annick, JENKINS Alan***
Management Accounting Change and the Construction of Coherence in Organisations: a Case Study

04004 ***CHARLETY Patricia, FAGART Marie-Cécile, SOUAM Saïd***
Real Market Concentration through Partial Acquisitions

04005 ***CHOFFRAY Jean-Marie***
La révolution Internet

04006 ***BARONI Michel, BARTHELEMY Fabrice, MOKRANE Mahdi***
The Paris Residential Market: Driving Factors and Market Behaviour 1973-2001

04007 ***BARONI Michel, BARTHELEMY Fabrice, MOKRANE Mahdi***
Physical Real Estate: A Paris Repeat Sales Residential Index

04008 ***BESANCENOT Damien, VRANCEANU Radu***
The Information Limit to Honest Managerial Behavior

04009 ***BIZET Bernard***
Public Property Privatization in France

04010 ***BIZET Bernard***
Real Estate Taxation and Local Tax Policies in France

04011 ***CONTENSOU François***
Legal Profit-Sharing: Shifting the Tax Burden in a Dual Economy

04012 ***CHAU Minh, CONTENSOU François***
Profit-Sharing as Tax Saving and Incentive Device

04013 ***REZZOUK Med***
Cartels globaux, riposte américaine. L'ère Empagran ?

2005

05001 ***VRANCEANU Radu***
The Ethical Dimension of Economic Choices

05002 ***BARONI Michel, BARTHELEMY Fabrice, MOKRANE Mahdi***
A PCA Factor Repeat Sales Index (1973-2001) to Forecast Apartment Prices in Paris (France)

05003 ***ALFANDARI Laurent***
Improved Approximation of the General Soft-Capacitated Facility Location Problem

05004 ***JENKINS Alan***
Performance Appraisal Research: A Critical Review of Work on "the Social Context and Politics of Appraisal"

05005 ***BESANCENOT Damien, VRANCEANU Radu***
Socially Efficient Managerial Dishonesty

05006 ***BOARI Mircea***
Biology & Political Science. Foundational Issues of Political Biology

05007 ***BIBARD Laurent***
Biologie et politique

05008 ***BESANCENOT Damien, VRANCEANU Radu***
Le financement public du secteur de la défense, une source d'inefficacité ?

2006

06001 ***CAZAVAN-JENY Anne, JEANJEAN Thomas***
Levels of Voluntary Disclosure in IPO prospectuses: An Empirical Analysis

06002 ***BARONI Michel, BARTHELEMY Fabrice, MOKRANE Mahdi***
Monte Carlo Simulations versus DCF in Real Estate Portfolio Valuation

06003 ***BESANCENOT Damien, VRANCEANU Radu***
Can Incentives for Research Harm Research? A Business Schools Tale

06004 ***FOURCANS André, VRANCEANU Radu***
Is the ECB so Special? A Qualitative and Quantitative Analysis

06005 ***NAIDITCH Claire, VRANCEANU Radu***
Transferts des migrants et offre de travail dans un modèle de signalisation

06006 ***MOTTIS Nicolas***
Bologna: Far from a Model, Just a Process for a While…

06007 ***LAMBERT Brice***
Ambiance Factors, Emotions and Web User Behavior: A Model Integrating and Affective and Symbolical Approach

06008 ***BATISTA Catia, POTIN Jacques***
Stages of Diversification and Capital Accumulation in an Heckscher-Ohlin World, 1975-1995

06009 ***TARONDEAU Jean-Claude***
Strategy and Organization Improving Organizational Learning

06010 ***TIXIER Daniel***
Teaching Management of Market Driven Business Units Using Internet Based Business Games

06011 ***COEURDACIER Nicolas***
Do Trade Costs in Goods Market Lead to Home Bias in Equities?

06012 ***AVIAT Antonin, COEURDACIER Nicolas***
The Geography of Trade in Goods and Asset Holdings

06013 ***COEURDACIER Nicolas, GUIBAUD Stéphane***
International Portfolio Diversification Is Better Than You Think

06014 ***COEURDACIER Nicolas, GUIBAUD Stéphane***
A Dynamic Equilibrium Model of Imperfectly Integrated Financial Markets

06015 ***DUAN Jin-Chuan, FULOP Andras***
Estimating the Structural Credit Risk Model When Equity Prices Are Contaminated by Trading Noises

06016 ***FULOP Andras***
Feedback Effects of Rating Downgrades

06017 ***LESCOURRET Laurence, ROBERT Christian Y.***
Preferencing, Internalization and Inventory Position

06018 ***BOURGUIGNON Annick, SAULPIC Olivier, ZARLOWSKI Philippe***
Management Accounting Change in the Public Sector: A French Case Study and a New Institutionalist Perspective

06019 ***de BEAUFORT Viviane***
One Share – One Vote, le nouveau Saint Graal ?

06020 ***COEURDACIER Nicolas, MARTIN Philippe***
The Geography of Asset Trade and the Euro: Insiders and Outsiders

06021 ***BESANCENOT Damien, HUYNH Kim, VRANCEANU Radu***
The "Read or Write" Dilemma in Academic Production: A European Perspective

2007

07001 ***NAIDITCH Claire, VRANCEANU Radu***
International Remittances and Residents' Labour Supply in a Signaling Model

07002 ***VIENS G., LEVESQUE K., CHAHWAKILIAN P., EL HASNAOUI A., GAUDILLAT A., NICOL G., CROUZIER C.***
Évolution comparée de la consommation de médicaments dans 5 pays européens entre 2000 et 2004 : analyse de 7 classes pharmaco-thérapeutiques

07003 ***de BEAUFORT Viviane***
La création d'entreprise au féminin dans le monde occidental

07004 ***BOARI Mircea***
Rationalizing the Irrational. The Principle of Relative Maximization from Sociobiology to Economics and Its Implications for Ethics

07005 ***BIBARD Laurent***
Sexualités et mondialisation

07006 ***VRANCEANU Radu***
The Moral Layer of Contemporary Economics: A Virtue Ethics Perspective

07007 ***LORINO Philippe***
Stylistic Creativity in the Utilization of Management Tools

07008 ***BARONI Michel, BARTHELEMY Fabrice, MOKRANE Mahdi***
Optimal Holding Period for a Real Estate Portfolio

07009 ***de BEAUFORT Viviane***
One Share - One Vote, the New Holy Graal?

07010 ***DEMEESTERE René***
L'analyse des coûts : public ou privé ?

07011 TIXIER Maud
Appreciation of the Sustainability of the Tourism Industry in Cyprus

07012 LORINO Philippe
Competence-based Competence Management: a Pragmatic and Interpretive Approach. The Case of a Telecommunications Company

07013 LORINO Philippe
Process Based Management and the Central Role of Dialogical Collective Activity in Organizational Learning. The Case of Work Safety in the Building Industry

07014 LORINO Philippe
The Instrumental Genesis of Collective Activity. The Case of an ERP Implementation in a Large Electricity Producer

07015 LORINO Philippe, GEHRKE Ingmar
Coupling Performance Measurement and Collective Activity: The Semiotic Function of Management Systems. A Case Study

07016 SALLEZ Alain
Urbaphobie et désir d'urbain, au péril de la ville

07017 de CARLO Laurence
The Classroom as a Potential Space - Teaching Negotiation through Paradox

07018 QU Chengxin
The Impact of Functionality Thresholds on Market Segmentation and Product Differentiation